Gregory Clunies Sunday, October 15, 2017

p0-Exploring Weather Trends

Global yearly temperature trends were compared against local yearly temperature trends for 3 different cities: Toronto, Canada, Seattle, USA, and Kingston, Jamaica. The acquisition of data, processing of data, and a brief analysis of the data are presented below.

Data Acquisition

Data was retrieved from the Udacity SQL workspace provided using the following commands:

```
/* To call list of cities */
SELECT *
FROM city_list

/* Also need to pull global data: */
SELECT *
FROM global_data

/* For Toronto, Canada data: */
SELECT *
FROM city_data
WHERE city IN ('Toronto');

/* For Seattle, USA data: */
SELECT *
FROM city_data
WHERE city IN ('Seattle');

/* For Kingston, Jamaica data: */
SELECT *
FROM city_data
WHERE city IN ('Kingston') AND country IN ('Jamaica');
```

Note that the AND operator was required for the call of the Kingston, Jamaica data as there were multiple cities named Kingston within the database (e.g., Kingston, Canada).

Data Processing

The data retrieved using the above SQL commands were exported in CSV format and loaded into excel where they were compiled to form a single table for analysis. After visual inspection of the data, it was found that the global and city data did not all span the same domain in time (i.e., Toronto time domain = 1743 to 2013, Seattle Time domain = 1828 to 2013, etc.). As a

result of this observation, datasets were trimmed to a time span from year 1828 to 2013 to ensure data was present for each year in the analysis. After trimming the data to a consistent time domain across all locations and the global data, erroneous values (blank cell in this case) were removed from the dataset. The identification of erroneous cells was done using the ISBLANK() function in Excel combined with conditional formatting to quickly identify a year in which any station or the global data had a blank value. Rows in which the ISBLANK() check returned the value TRUE were deleted to This method is shown if Figure 1 below.

I10 =OR(ISBLANK(B10),ISBLANK(D10),ISBLANK(F10),ISBLANK(H10))

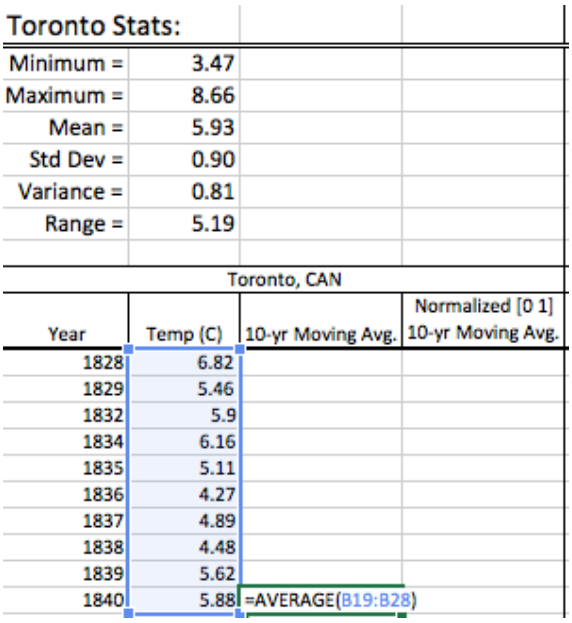
	A	B	C	D	E	F	G	H	I
1	**Notes:**								
2	- All data before 1828 and after 2013 was removed so that comparisons and/calcs are over a period for which data is avai								
3	Toronto, Seatte, Kingston, and Global temps.								
4	- Blank data entries were found using ISBLANK() function in EXCEL and conditional formatting (for ease of use),								
5	these rows will be excluded from furtehr analysis (next worksheet).								
6									TRUE = cell IS
7									FALSE = cell IS
8	Toronto, CAN		Seattle, USA		Kinsgton, JAM		Global		ISBLANK ?
9	Year	Temp (C)	Year	Temp (C)	Year	Temp (C)	Year	Temp (C)	
10	1828	6.82	1828	7.13	1828	26.94	1828	8.17	FALSE
11	1829	5.46	1829	6.8	1829	26.23	1829	7.94	FALSE
12	1830	6.23	1830		1830	26.92	1830	8.52	TRUE
13	1831	4.96	1831		1831	25.98	1831	7.64	TRUE
14	1832	5.9	1832	3.52	1832	25.48	1832	7.45	FALSE
15	1833	5.88	1833	7.48	1833		1833	8.01	TRUE
16	1834	6.16	1834	7.1	1834	23.93	1834	8.15	FALSE
17	1835	5.11	1835	5.58	1835	25.8	1835	7.39	FALSE
18	1836	4.27	1836	6.62	1836	25.78	1836	7.7	FALSE
19	1837	4.89	1837	6.37	1837	26.02	1837	7.38	FALSE
20	1838	4.48	1838	6.59	1838	25.85	1838	7.51	FALSE
21	1839	5.62	1839	7.3	1839	26.11	1839	7.63	FALSE
22	1840	5.88	1840	6.69	1840	26.13	1840	7.8	FALSE
23	1841	5.33	1841	6.81	1841	25.84	1841	7.69	FALSE
24	1842	5.75	1842	6.88	1842	26.24	1842	8.02	FALSE
25	1843	4.81	1843	6.55	1843	26.1	1843	8.17	FALSE
26	1844	5.8	1844	6.41	1844	25.64	1844	7.65	FALSE
27	1845	5.81	1845	6.88	1845	25.69	1845	7.85	FALSE
28	1846	6.47	1846		1846	26.5	1846	8.55	TRUE
29	1847	5.11	1847	8.99	1847	26.25	1847	8.09	FALSE
30	1848	5.68	1848	6.6	1848	26.22	1848	7.98	FALSE

Deleted Rows with Blank Values

Analysis

To help remove the variability of the datasets and to aid in comparison, 10-yr moving averages were calculated in Excel and shown below in Figure 2 (left).

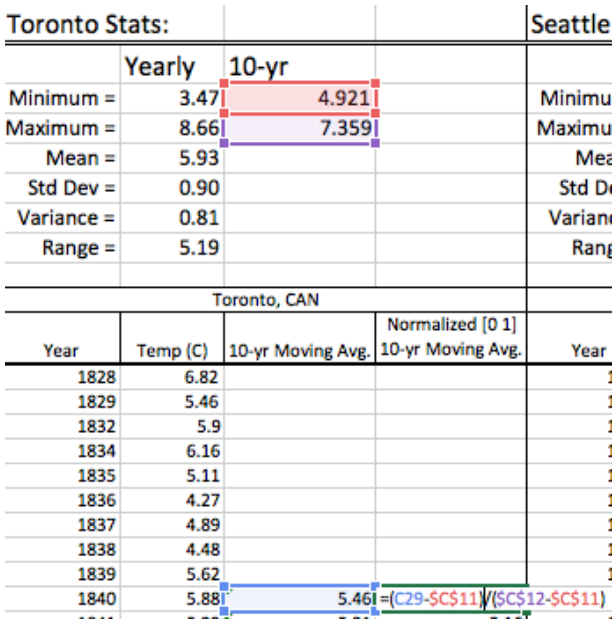
Figure 2. Example calculations for 10-yr moving average and normalization of data [0 1].



Toronto Stats:	
Minimum =	3.47
Maximum =	8.66
Mean =	5.93
Std Dev =	0.90
Variance =	0.81
Range =	5.19

Toronto, CAN

Year	Temp (C)	10-yr Moving Avg.	Normalized [0 1] 10-yr Moving Avg.
1828	6.82		
1829	5.46		
1832	5.9		
1834	6.16		
1835	5.11		
1836	4.27		
1837	4.89		
1838	4.48		
1839	5.62		
1840	5.88	=AVERAGE(B19:B28)	



Toronto Stats:	Yearly	10-yr	Seattle
Minimum =	3.47	4.921	Minimu
Maximum =	8.66	7.359	Maximu
Mean =	5.93		Mea
Std Dev =	0.90		Std D
Variance =	0.81		Varian
Range =	5.19		Rang

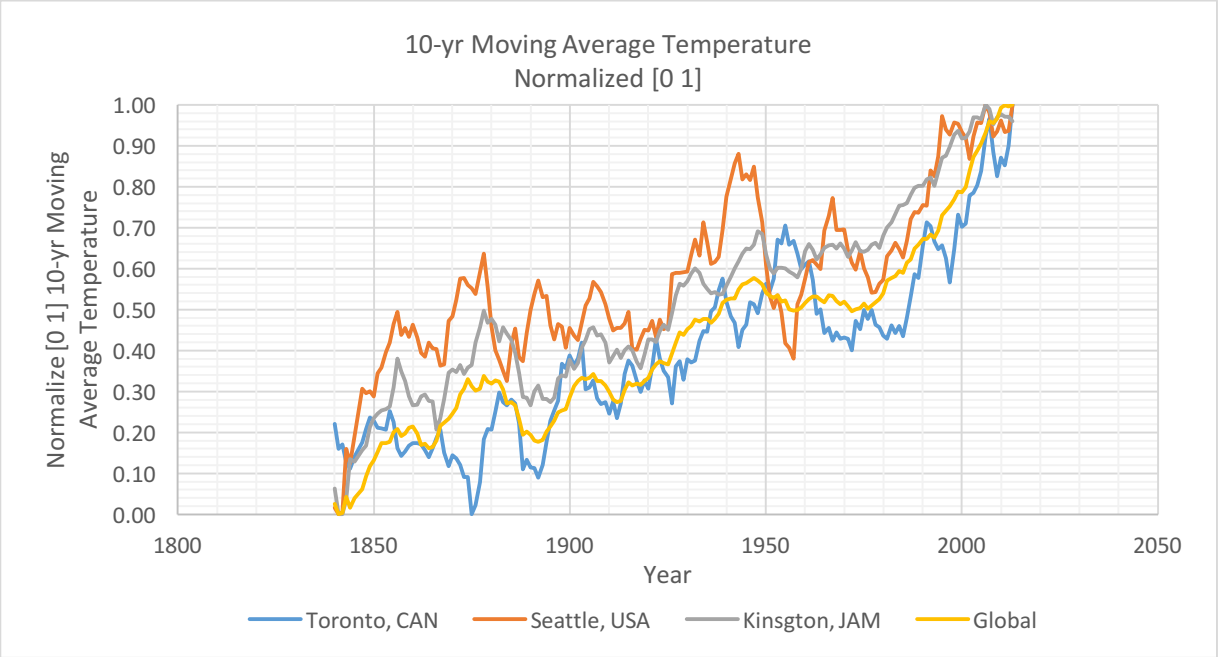
Toronto, CAN

Year	Temp (C)	10-yr Moving Avg.	Normalized [0 1] 10-yr Moving Avg.	Year
1828	6.82			1
1829	5.46			1
1832	5.9			1
1834	6.16			1
1835	5.11			1
1836	4.27			1
1837	4.89			1
1838	4.48			1
1839	5.62			1
1840	5.88	5.46	=(C29-C11)/(C12-C11)	

Additionally, the dataset for each city and the global data was normalized so that the data ranged from 0 to 1. This provides a sense of how the temperature for each city has progressed from 1828 to 2013 across their respective ranges. An example of this normalization of the data is also shown in Figure 2 (right). The resulting datasets were plotted on the same axes for comparison and are shown in Figure 3. Based on review of the data the following general conclusions can be made:

1. All three cities have experienced temperature increase as has the globe as a whole.
2. The global temperature increase has been much more steady than the temperature increase experienced by the individual cities, which have experienced much higher year-to-year variation. This is still evident in the 10-yr moving averages.
3. Jamaica is the only city who's 10-yr moving average temperature in 2013 was not within 2% of the maximum observed from 1828 to 2013.
4. Relative to the overall trend of temperature increase from 1828 to 2013, Seattle had a decrease in temperature between 1943 and 1957, losing roughly 50% of its ultimate temperature increase from 1828 to 2013.

Figure 3. 10-yr moving averages normalized to 0 to 1.



Correlation scores for the 10-yr moving averages across all the cities and the global temperature were calculated using the CORREL() function in excel. It was found that Jamaica, Kingston had the highest correlation score with the global temperature at 0.98. This is also supported visually in figure 3 as increases/decreases in global temperature are often mirrored by increases/decreases in the temperature in Kingston of similar magnitude (although not perfectly matched). A summary of correlation scores is shown in Table 1 below.

Table 1. Correlation Scores: 10-yr Moving Avg. Temperatures

	Toronto, CAN	*Seattle, USA*	*Kingston, JAM*	*Global*
Toronto, CAN	1.00			
Seattle, USA	0.74	1.00		
Kingston, JAM	0.89	0.90	1.00	
Global	0.92	0.90	0.98	1.00

In order to get some idea of what the 10-yr moving average temperature of Kingston might be based solely from an observed global temperature, global 10-yr moving average temperature values of were plotted against their respective values for Kingston. The relationship between Kingston and global 10-yr moving average temperatures are largely linear. Using Excel, a linear

regression was performed for which the line of best fit is shown in figure 4 below along with its corresponding equation.

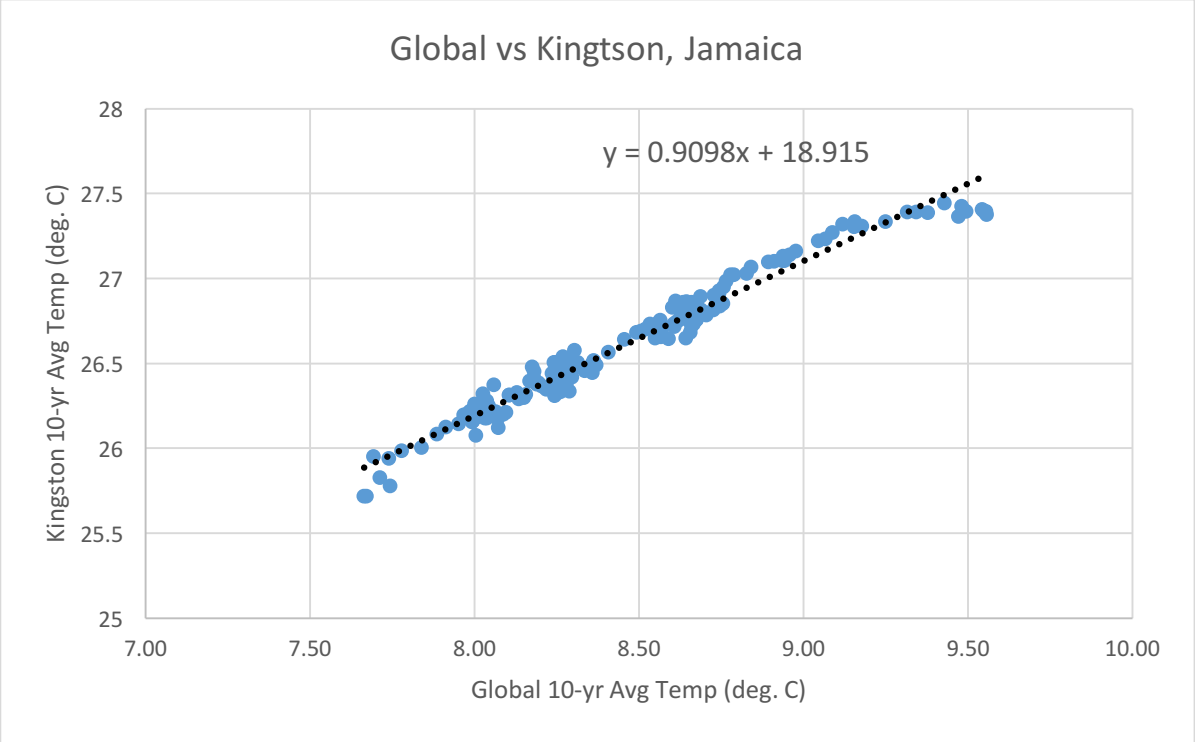


Figure 4. 10-yr moving average temperatures for Kingston, Jamaica plotted against global 10-yr moving average temperatures.

Based on the observed data, the equation to estimate the 10-yr moving average temperature for Kingston, Jamaica based on an observed global 10-yr moving average temperature is:

$y = 0.9098x + 18.915$

where, y = Estimated Kingston, Jamaica 10-yr moving average temperature
x = Observed Global 10-yr moving average temperature